# A panoramic view of bacterial transcription

*Jean-Francois Tomb*

The availability of whole genome sequences has ushered in a new era in the biological sciences. For the first time in human history, all (or most) of the genes of an organism can be represented. Translating this into useful functional information represents one of the greatest challenges to the research community today. In this regard, rapidly developing DNA array technology has already provided some important insights into gene function in yeast, murine, and human cells. Now, Jan Mous and colleagues[1] present one of the first reports using high-density oligonucleotide probe arrays to monitor gene expression in bacteria. Preliminary results obtained for 100 *Streptococcus pneumoniae* genes are extremely encouraging and demonstrate the applicability of oligonucleotide microarrays to global gene expression monitoring of a bacterial genome.

To date, complete genetic blueprints have been obtained for 14 organisms, spanning the three domains of life—Eukarya, Archea, and bacteria. The genomic sequence of thirty or more additional microbes will be known in the next three years[2]. Although this flood of sequence information has led to a great increase in gene discovery, it is also clear that at least 40% of the identified putative gene products in each genome remain poorly characterized, and many are potentially unique to a particular species. These data underscore how little is actually known about microbial physiology and diversity and genome plasticity. Elucidating the role of all proteins in a genome constitutes an enormous challenge to both experimental and computer scientists. It represents, however, a unique opportunity to discover new enzymes and metabolic pathways, new vaccine candidates and drug targets, and define new structure–function relationships.

Genomics has emerged from the process of developing high-throughput sequencing and from establishing the infrastructure required to track, analyze, and comprehensively represent sequence and annotation data. On the basis of the informational content of the complete DNA sequence of a genome, computational genomics formulates hypotheses regarding the biology of an organism, including its physiology, genetics, virulence, and evolution. To test these hypotheses and uncover new biological paradigms, experimental genomics applies genome-wide (global) approaches aimed at measuring levels of gene expression, identifying essential genes, and determining functional requirements for the various biological processes (Fig. 1).

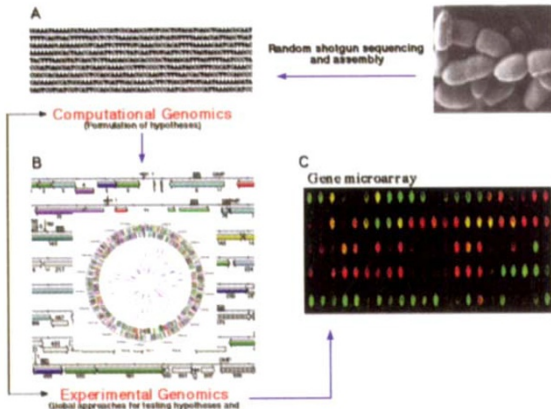


**Figure 1. A schematic view of a microbial genome project. The DNA is isolated from a particular organism, purified, mechanically sheared and the fragments cloned and sequenced. A. Several thousand sequences are assembled and the gaps are closed to produce a single contiguous sequence. B. The sequence is annotated and hypotheses regarding the gene content, metabolic and transport capacity, the genetic and regulatory repertoire, and the evolution of the organism are advanced. C. To gain a dynamic view of the genome and verify these hypotheses, several genome-wide experimental approaches are used (e.g., microarrays).**

Improved methods of analysis and annotation of microbial genomes are being developed (e.g., see ref. 3). Strategies for global expression and functional analyses are also being refined and include serial analysis of gene expression (SAGE)[4], DNA and oligonucleotide microarrays[1,5], two-dimensional gel electrophoresis/mass spectrometry[6], genetic footprinting[7], and mutagenesis.

In the approach adopted by Mous and colleagues, each gene in the genome is represented by a set of overlapping oligomers, which are synthesized in situ using a light-directed chemical synthesis process employed in the semiconductor industry[5] (see pp. 27 and 40 of this issue). Other array formats are also available; for example, high-density spotting on a microscope slide of DNA fragments produced by polymerase chain reaction (PCR) amplification. Whatever the method of choice, the chips are probed with fluorescently labeled or biotinylated cDNA obtained by reverse transcribing RNA. (In the case of prokaryotes, total RNA, containing ~95% ribosomal RNA, is used as a substrate[5].) The signals are then captured using confocal microscopy.

Mous et al. demonstrate that the signals obtained from probing oligomer microarrays are specific (no cross hybridization), are proportional to the amount of input RNA, and correlate well with results obtained by northern analyses. Furthermore, they show that a twofold change in expression is detectable and a sensitivity of one to five transcripts per cell is achievable. These results are highly reproducible, with a variation of expression up to 25%. The authors recognize, however, the current limitation of the approach for monitoring gene expression in vivo: Better enzymology and labeling efficiencies are required.

The authors' approach has clear advantages over other less reproducible methods, such as differential display of amplified RNA on sequencing gels or low-throughput approaches such as northern analyses. And, on the basis of their work, Mous et al. caution that other array formats, such as cDNA microarrays, may be unsuitable for prokaryote genome analysis because of high background observed when using total RNA as labeled probe. While valuable, this cautionary note contradicts results obtained by Blattner's group[8] using high-density arrays on nylon filters.

Finally, although the expression data obtained by these high-throughput technologies are of immense value, they provide only a partial picture of the regulatory repertoire of an organism. Translational regulation and post-translational modification of proteins represent aspects of functional control beyond the scope of this methodology. Nevertheless, the panoramic view of transcription provided by this approach is sufficiently breathtaking to make us eagerly await its application to further microbial genomes.

*Jean-Francois Tomb is associate investigator, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850 (hanna@tigr.org).*

1. de Saizieu, A. et al. 1998. *Nature Biotechnology* **16:**45–48.
2. http://www.tigr.org
2. Sonnhammer, E.L.L. et al. 1997. *Proteins* **28:**405–420.
3. Velculescu, V.E. et al. 1995. *Science* **270:**484–487.
4. Schena, M. et al. 1995 *Science* **270:**467–470 and http://www.affymetrix.com
5. Link, A.J. et al. 1997. *Electophoresis* **18:**1314–1334.
6. Smith, V. et al. 1996. *Science* **274:**2069–2074.
7. Chuang, S.E. et al. 1993. *J. Bacteriol.* **175:**2026–2036.